# BELIEF FUNCTIONS AND RANDOM SETS

A LECTURE DELIVERED AT THE

SPRING SCHOOL ON BELIEF FUNCTIONS

*CARTHAGE, TUNISIA, MAY 2013*

By

HUNG T. NGUYEN

# CONTENTS

**A. Belief functions on finite domains**
1. Where should we begin?
2. Characterization of beliefs
3. Capacity functionals
4. Combination of evidence
**B. Belief functions on infinite domains**
1. Random closed sets
2. Capacity functionals

**References**

## Outline of the lecture

This lecture is tutorial in nature and aims at laying down the foundations for studying belief functions within the rigorous framework of random sets in probability theory.

While the main focus of the lecture is on *belief functions on infinite domains*, where we present, in its elementary form, the theory of *random closed sets* as foundations, we briefly recall first the familiar setting of belief functions on finite domains, not only to make the transition to infinite domains, but also to discuss several important aspects in uncertainty modeling and inference with belief functions, including the problem of combination of dependent pieces of evidence via the notion of copulas.

While the actual presentation of the lecture, in view of time limit, is much more concise, the material detailed here will serve as a complete set of lecture notes.

# A. BELIEF FUNCTIONS ON FINITE DOMAINS

## A.1. WHERE SHOULD WE BEGIN ?

Without starting like the beginning of any topic with a prehistory, then a history, we simply proceed, using the context of *coarse data analysis* in statistics, to reach the notion of "degree of belief" which is the starting point of the theory of belief functions.

Discovering knowledge using empirical data is the norm of scientific studies. However, there are situations where the variable of interest $X$ is latent, i.e., not observable. A familiar case is that the values of $X$ cannot be "measured" or observed with accuracy, instead, some forms of approximate values of $X$ can be assessed, such as each value of $X$, say, in an experiment, is only known to be in some "interval" or set $S$, where the set $S$ is observed. The description of such a situation is based on the concept of *selectors.* Specifically, let $X, S$ be two random elements, defined on the same probability space $(\Omega, \mathcal{A}, P)$ with values in the spaces $U$ (e.g., finite) and $2^U$, respectively. While $X : \Omega \to U$ is a random variable, $S : \Omega \to 2^U$ is a random set since it takes subsets of $U$ as values (or outcomes).

Saying that $X$ is an almost sure (a.s.) selector of $S$, we mean $P(X \in S) = 1$, that is, with probability one, $X(\omega) \in S(\omega)$. In this case, the data on $X$ given by its "coarsening" $S$ is referred to as "coarse data", i.e., data with low quality.

Since $P(X \in S) = 1$, the random set $S$ is a *nonempty random set,* i.e., $P(S = \varnothing) = 0$.

*Remark on coarsening schemes*

It is a remarkable intelligent behavior of humans to be able to use coarsening schemes in their everyday decision-making processes. For example, when you do not know the exact age of someone, you describe your perception using a coarsening of the range $[0, 120]$ of the form $[0, 20], (20, 40], (40, 70], (70, 120]$ (i.e., a partition of $[0, 120]$), to say, e,g., the person's age is between 20 and 40; or more "sophisticatedly", using a fuzzy coarsening scheme "young", "middle-age", "old" (just like in fuzzy rules in control). A coarsening scheme of a domain is a transformation of the "numerical scale" into a set-valued scale. Each element of a coarsening scheme contains the unobservable value

of the variable of interest (e.g., age) that you "believe". Thus, underlying a coarsening scheme $S$ is some selector $X$.

Back to our above situation of coarse data: Recall that, related to a random variable $X$, a subset $A \subseteq U$ is an event in $\mathcal{A}$, in the sense that $X^{-1}(A) \in \mathcal{A}$, its probability of occurence is $P(X \in A)$. We say that $A$ occurs when $X(\omega) \in A$. Now, we cannot observe $X(\omega)$ and but can observe only $S(\omega)$. There are three cases: If $S(\omega) \subseteq A$, then we are sure that $A$ occurs, since $X(\omega) \in S(\omega)$; if $S(\omega) \cap A = \varnothing$, then we are sure that $A$ do not occur; but if $S(\omega) \cap A \neq \varnothing$ with $S(\omega) \cap A^c \neq \varnothing$ it is not clear what we can say, since while $X(\omega) \in S(\omega)$, it may happen that $X(\omega) \in S(\omega) \cap A^c$. From a *pestimistic view point*, we could "believe" that $A$ occurs when $S(\omega) \subseteq A$, i.e., quantify our "degree of belief" on the occurence of $A$ by $P(S \subseteq A)$ (say, from an empirical viewpoint, use $\frac{1}{n}\#\{i : S_i \subseteq A\}$). Note that the degree of belief (in the occurence of a set $A$) is viewed as an approximation to its probability of occurence $P(X \in A)$, and is less than this probability: since $\{\omega : S(\omega) \subseteq A\} \subseteq \{\omega : X(\omega) \in A\}$, $P(S \subseteq A) \leq P(X \in A)$.

So let's consider a *quantification of beliefs,* in the above context of coarse data, as the set function $F : 2^U \to [0, 1]$, defined by

$$F(A) = P(S \subseteq A)$$

## A.2. CHARACTERIZATION OF BELIEFS

Let $U$ be a finite set. A random set is a random element $S$, defined on a probability space $(\Omega, \mathcal{A}, P)$ and taking values in the power set $2^U$ of $U$. Specifically, the $\sigma-$field of $2^U$ is taken to be $2^{2^U}$, and $S$ is $\mathcal{A} - 2^{2^U} -$ measurable, i.e., for any $\mathbb{A} \in 2^{2^U}$, $S^{-1}(\mathbb{A}) \in \mathcal{A}$. Since $2^{2^U}$ is finite, it suffices that, for any $A \in 2^U$, $S^{-1}(\{A\}) = \{\omega : S(\omega) = A\} \in \mathcal{A}$, since $S^{-1}(\mathbb{A}) = \cup_{A \in \mathbb{A}} S^{-1}(\{A\})$. The random set $S$ is said to be nonempty if $P(S = \varnothing) = 0$.

From the above motivation, consider the quantification of beliefs by the set function $F : 2^U \to [0, 1]$, defined by $F(A) = P(S \subseteq A)$,where $S$ is a nonempty random set on $U$. We get the following basic properties of the set function $F(.)$

**Theorem 1**

The set function $F(.) : 2^U \to [0, 1]$, where $F(A) = P(S \subseteq A)$, satisfies the following properties

(i) $F(\varnothing) = 0$

(ii) $F(U) = 1$

(iii) For any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $2^U$,

$$F(\cup_{j=1}^{k} A_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

where $|I|$ denotes the cardinality of the set $I$.

**Proof**

Clearly, $F(\varnothing) = P(S = \varnothing) = 0$, and $F(U) = P(S \subseteq U) = 1$.

For (iii), we proceed as follows. Let $k \geq 2$, and $A_1, A_2, ..., A_k$ be given. For $B \subseteq U$, let

$$J(B) = \{i = 1, 2, ..., k \quad \text{such that} \quad B \subseteq A_i\}$$

i.e., $B \subseteq \cap_{i \in J(B)} A_i$.

Let $f(A) = P(S = A)$. Then

$$F(\cup_{j=1}^{k} A_j) = P(S \subseteq \cup_{j=1}^{k} A_j) = \sum_{B \subseteq \cup_{j=1}^{k} A_j} f(B) \geq \sum_{B \subseteq U, J(B) \neq \varnothing} f(B)$$

by noting that $\{B : B \subseteq \cup_{j=1}^{k} A_j\} \supseteq \{B \subseteq U : J(B) \neq \varnothing\}$.

Now, since for $J(B) \neq \varnothing$, $\sum_{\varnothing \neq I \subseteq J(B)} (-1)^{|I|+1} = 1$ (in fact, for any nonempty and finite set $A$, $\sum_{\varnothing \neq B \subseteq A} (-1)^{|B|+1} = 1$), we can write

$$\sum_{B \subseteq U, J(B) \neq \varnothing} f(B) = \sum_{B \subseteq U, J(B) \neq \varnothing} [\sum_{\varnothing \neq I \subseteq J(B)} (-1)^{|I|+1}] f(B)$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} \sum_{B \subseteq U, J(B) \supseteq I} f(B) = \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} \sum_{B \subseteq \cap_{i \in I} A_i} f(B)$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

by interchanging the order of the summations.

*Remarks*

a) The property (iii) is referred to as the *monotonicity of infinite order*. If (iii) is satisfied for $k = 2$, we say that $F$ is monotone of order 2. Since $F$

is monotone of infinite order with $F(\varnothing) = 0$, it is monotone increasing: $A \subseteq B \Longrightarrow F(A) \leq F(B)$. This follows from the general fact that if $F : 2^U \to \mathbb{R}$, monotone of order 2, then $F$ is monotone if and only if $F(\varnothing)$ is the minimum value of $F(.)$. Indeed, if $F$ is monotone, then clearly $F(\varnothing)$ is its minimum value. Conversely, suppose $F(\varnothing)$ is its minimum value. Let $A \subseteq B$. Then $B = A \cup C$ with $A \cap C = \varnothing$. By $2-$ monotonicity, $F(B) = F(A \cup C) \geq F(A) + F(C) - F(A \cap C) = F(A) + F(C) - F(\varnothing)$. Since $F(C) \geq F(\varnothing)$, it follows that $F(C) - F(\varnothing) \geq 0$, and hence $F(B) \geq F(A)$.

b) The property (iii) generalizes Henri Poincare' s equality for probability measures. Recall that if $Q$ is a probability measure on $2^U$, then for $A, B$ in $2^U$, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

or more generally, for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $2^U$,

$$Q(\cup_{j=1}^{k} A_j) = \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} Q(\cap_{i \in I} A_i)$$

The above property of monotonicity of infinite order is simply a weakening of Henri Poincare's equality (to an inequality). A set function which is monotone of infinite order is not necessarily additive. Thus, if we view such a set function as a "measure of uncertainty", it is a nonadditive measure. Perhaps, this is the reason why we see statements such as "there are many situations involving uncertainty which cannot be represented properly within the classical probability framework". As we will see, it all depends on how we look at things. Put it differently, "the point of view is everything"!

The upshot is that the above basic properties of beliefs in Theorem 1 characterize "distribution functions" of random sets, where by the *distribution function* of a random set $S$, we mean the set function $F(A) = P(S \subseteq A)$, by analogy with a distribution function of a real-valued random variable, in which the total order $\leq$ on the reals is replaced by the partial order $\subseteq$ of set inclusion. It is clear that for *finite random sets*, their distributions determine completely their probability laws (measures) on $2^{2^U}$, via their probability *density functions* on $2^U$. Indeed, the probability law of the random element $S$ is the probability measure $P_S$ on $2^{2^U}$ which can be written as $P_S(\mathbb{A}) = P(S \in \mathbb{A}) = \sum_{A \in \mathbb{A}} P(S = A)$, where $f : 2^U \to [0, 1]$, defined as $f(A) = P(S = A)$ is a bona fide probability density function on $2^U$, i.e., it is a nonnnegative function with $\sum_{A \subseteq U} f(A) = 1$.

**Theorem 2**
Let $F : 2^U \to [0,1]$ such that
(i) $F(\varnothing) = 0$
(ii) $F(U) = 1$
(iii) For any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $2^U$,

$$F(\cup_{j=1}^{k} A_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

then $F(.)$ is the distribution of some nonempty random set $S$.

**Proof**
To show the existence of a random element $S$ taking values in $2^U$ (finite) having the given set function $F(.)$ as its distribution, it suffices to show that there is a probability density function $f : 2^U \to [0,1]$ such that $F(A) = P(S \subseteq A) = \sum_{B \subseteq A} f(B)$. The following will do

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

Of course, we need to show that this function $f(.)$ on $2^U$ is indeed a probability density function.

Since $F(\varnothing) = 0$, it follows that $f(\varnothing) = 0$, and by construction, $f(\{u\}) = F(\{u\}) \geq 0$. For $A = \{u, v\}$, we have,

$$f(\{u,v\}) = \sum_{B \subseteq \{u,v\}} (-1)^{|A \setminus B|} F(B) = F(\{u,v\}) - F(\{u\}) - F(\{v\})$$

But, by (iii), with $k = 2$, $A_1 = \{u\}$, $A_2 = \{v\}$, noting that $A_1 \cap A_2 = \varnothing$, we have

$$F(\{u,v\}) \geq F(\{u\} + F(\{v\})$$

and hence $f(\{u,v\}) \geq 0$. More generally,

For $A \subseteq U$ with $|A| \geq 2$, say, $A = \{u_1, u_2, ..., u_n\}$, let $A_i = A \setminus \{u_i\}$. Then, by (iii) with $A = \cup_{i=1}^{n} A_i$, and by noting that $\cap_{i=1}^{n} A_i = \varnothing$,

$$f(A) = F(A) - \sum_{i=1}^{n} F(A_i) + \sum_{i<j} F(A_i \cap A_j) + .... + (-1)^{n-1} \sum_{i=1}^{n} F(\cap_{j \neq i} A_j) \geq 0$$

Thus, for any $A \in 2^U$, $f(A) \geq 0$.
That $F(A) = \sum_{B \subseteq A} f(B)$ and $\sum_{A \subseteq U} f(A) = 1$ will follow from

$$\sum_{B \subseteq A} f(B) = \sum_{B \subseteq A} [\sum_{C \subseteq B} (-1)^{|B \backslash C|} F(C)] = \sum_{C \subseteq B \subseteq A} (-1)^{|B \backslash C|} F(C)$$

which can be seen as follows. If $C = A$, then the last expression is $F(A)$. If $C \neq A$, then $A \backslash C$ has $2^{|A \backslash C|}$ subsets, so there are an even number of subsets $B$ with $C \subseteq B \subseteq A$, exactly half of which have an even number of elements. The half of the numbers $(-1)^{|B \backslash C|} F(C)$ are 1, and half are $-1$. Thus, for each $C \neq A$, $\sum_{C \subseteq B \subseteq A} (-1)^{|B \backslash C|} F(C) = 0$ with the summation taken over $B$. Therefore, $\sum_{B \subseteq A} f(B) = F(A)$. In particular, $\sum_{B \subseteq U} f(B) = F(U) = 1$.

*Remarks*
(a) *Random set representations of belief functions*
The set function $f : 2^U \rightarrow [0, 1]$, given by

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \backslash B|} F(B)$$

is a bona fide probability density function on $2^U$. As such, it is the density of some random set $S : (\Omega, \mathcal{A}, P) \rightarrow 2^U$ with $P(S = A) = f(A)$ for $A \in 2^U$. Any such random set is called a random set representation of the belief function $F$. Here is a *canonical* random set representation. If $f(A) > 0$, then we call the set $A \in 2^U$ a "focal element" of $F$ (or of $f$). Let $\mathcal{F}(F)$ be the set of focal elements of $F$, say, with $|\mathcal{F}(F)| = n$. Let $A_1, A_2, ..., A_n$ be a numeration of elements of $\mathcal{F}(F)$. Take $(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B}_1, \lambda(dx))$, where $\mathcal{B}_1$ is the Borel $\sigma-$field on $[0, 1]$ and $\lambda(dx)$ is the Lebesgue measure on $[0, 1]$. Partition $[0, 1]$ into $n$ intervals $I_i$ with length $\lambda(I_i) = f(A_i), i = 1, 2, ..., n$. Specifically,

$$I_1 = [0, f(A_1))$$

and for $2 \leq i \leq n - 1$,

$$I_i = [\sum_{j=1}^{i-1} f(A_j), \sum_{j=1}^{i} f(A_j))$$

and

$$I_n = [\sum_{j=1}^{n-1} f(A_j), 1]$$

Define $S : [0,1] \to \mathcal{F}(F)$ by $S(x) = A_i$ for $x \in I_i$, $i = 1, 2, ..., n$. Then, clearly,

$$P(S = A_i) = \lambda(x : S(x) = A_i) = \lambda(I_i) = f(A_i),\ i = 1, 2, ..., n$$

Thus, for $A \in 2^U$,

$$P(S \subseteq A) = \sum_{A_j \subseteq A} P(S = A_j) = \sum_{A_j \subseteq A} f(A_j) = F(A)$$

Since there are $n!$distinct orderings of the focal elements, there are $n!$ canonical random set representations for $F$. But all canonical random sets are equivalent in the sense that they all have the same distribution. Indeed, let $\sigma$ be a permutation on $\{1, 2, ..., n\}$, then $[0,1]$ is partitioned as

$$I_{\sigma(1)} = [0, f(A_{\sigma(1)}))$$

and for $2 \leq i \leq n-1$,

$$I_{\sigma(i)} = [\sum_{j=1}^{i-1} f(A_{\sigma(j)}), \sum_{j=1}^{i} f(A_{\sigma(j)}))$$

and

$$I_n = [\sum_{j=1}^{n-1} f(A_{\sigma(j)}), 1]$$

Let $S_\sigma : [0,1] \to \mathcal{F}(F)$ be defined as $S_\sigma(x) = A_{\sigma(i)}$ for $x \in I_{\sigma(i)}$. Then,

$$P(S_\sigma = A_{\sigma(i)}) = \lambda(I_{\sigma(i)}) = f(A_{\sigma(i)})$$

Thus, for $A \in \mathcal{F}(F)$, say $A = A_i = A_{\sigma(j)}$ where $\sigma(j) = i$, we have

$$P(S = A) = P(X = A_i) = f(A_i)$$

$$P(S_\sigma = A) = P(S_\sigma = A_{\sigma(j)}) = f(A_{\sigma(j)}) = f(A_i)$$

*Remark*

Of course, we can consider other types of random set reprentations such as discrete ones: For a numeration of the focal elements $\mathcal{F}(F)$, say, $A_i$, $i = 1, 2, ..., n$, take $\Omega = \{1, 2, ..., n\}$ and a probability measure on $\Omega$ defined, for $B \subseteq \Omega$, by $P(B) = \sum_{i \in B} f(A_i)$, and define $S : \Omega \to 2^U$ by $S(i) = A_i$, $i = 1, 2, ...n$. Then, for any $A \in 2^U$, $P(S = A) = f(A)$ for $A \in \mathcal{F}(F)$, and zero otherwise. Thus, for $A \in \mathcal{F}(F)$, $F(A) = P(S \subseteq A)$.

(b) **Axiomatic definition of belief functions.**

The conditions in the above Theorem 2 are taken as axioms to provide an *axiomatic definition of a belief function* in the literature. From this Theorem, we see that, whether or not there is "randomness around", the fact that we do have a bona fide probability density function $f$ on $2^U$ allows us to place the setting of belief functions completely within the standard framework of probability theory. Although the belief function $F(.)$, as a set function, is not additive, it does generate a probability measure at a higher level ($2^{2^U}$) namely $P_S(\mathbb{A}) = \sum_{A \in \mathbb{A}} f(A)$, just like distribution functions of random vectors.

(c) **Incidence algebras**

Why do we think about the function $f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$?

To be fair, we should evoke the sophisticated theory of *incidence algebra in combinatorics* (proposed by C. Rota in 1964 as a basis for combinatorial theory from a unified point of view). So let me say few words about it.

First of all, in the context of finite random sets, we are dealing with the (finite) set $2^U$ which is a partially ordered set (poset) with set inclusion $\subseteq$ as the partial order relation among subsets (and hence a special case of locally finite posets, where a poset $(X, \leq)$ is said to be locally finite if every "interval" $\{x : a \leq x \leq b\}$ is finite, such as the set of integers $\mathbb{Z}$).

In fact, we are dealing with real-valued functions defined on $2^U$ and $2^U \times 2^U$.

The set $\mathcal{V} = \{f : 2^U \to \mathbb{R}\}$, with pointwise addition and multiplication $r \in \mathbb{R}$, $rf(.)$, is a vector space over $\mathbb{R}$, with dimension $2^{|U|}$.

The set $\mathcal{W} = \{\alpha : (2^U)^{[2]} \to \mathbb{R}\}$, where

$$(2^U)^{[2]} = \{(A, B) : A \subseteq B \subseteq U\}$$

with pointwise addition and multiplication defined by

$$(\alpha * \beta)(A, B) = \sum_{A \subseteq Z \subseteq B} \alpha(A, Z)\beta(Z, B)$$

is called the *incidence algebra* of $U$ over the field $\mathbb{R}$. It is a ring with identity which is the function on $\mathcal{W}$ defined by

$$\delta(A,B) = \begin{cases} 1 & \text{for} \quad A = B \\ 0 & \text{for} \quad A \neq B \end{cases}$$

For an element $\alpha \in \mathcal{W}$ to have an inverse, it is necessary and sufficient that for all $A$, $\alpha(A,A) \neq 0$, and then its inverse is given inductively by

$$\alpha^{-1}(A,A) = \frac{1}{\alpha(A,A)}$$

$$\alpha^{-1}(A,B) = [\frac{-1}{\alpha(A,A)}] \sum_{A \subseteq Z \subseteq B} \alpha(A,Z)\alpha^{-1}(Z,B)$$

for $A \subseteq B$.

Elements that have an inverse are called *units*. There are two very special units in $\mathcal{W}$:

(1) The *Mobius function* $\mu(A,B) = (-1)^{|B \setminus A|}$ (for $A \subseteq B$, and 0 otherwise),

(2) The *Zeta function* $\xi(A,B) = 1$ (for $A \subseteq B$, and 0 otherwise).

which are inverses of each other, i.e., $\mu * \xi = \xi * \mu = \delta$

There is a natural operation on the elements of the vector space $\mathcal{V}$ by the elements of the incidence algebra $\mathcal{W}$: For $f \in \mathcal{V}$, and $\alpha \in \mathcal{W}$, define $f * \alpha : 2^U \to \mathbb{R}$ by

$$(f * \alpha)(A) = \sum_{Z \subseteq A} f(A)\alpha(Z,A)$$

In particular,

$$(f * \mu)(A) = \sum_{Z \subseteq A} f(A)\mu(Z,A) = \sum_{Z \subseteq A} (-1)^{|A \setminus Z|} f(Z)$$

is called the *Mobius inverse* of $f$ or the Mobius inversion of $f$.

Thus, the probability density function (or "mass" function)

$$f(A) = (F * \mu)(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

is the Mobius inverse of the belief function $F$.

The maps $f \in \mathcal{V} \to f * \mu \in \mathcal{V}$ and $f \to f * \xi$ are one to one maps from $\mathcal{V}$ onto $\mathcal{V}$, and are inverses of one another. Specifically, let $\varphi(f) = f * \mu$, and $\psi(f) = f * \xi$, then $\varphi \circ \psi = \psi \circ \varphi = id$. Indeed,

$$(\varphi \circ \psi)(f)(A) = \varphi[\psi(f)](A) = \varphi[f * \xi](A) = [(f * \xi) * \mu](A)$$

$$= [f * (\xi * \mu)](A) = (f * \delta)(A) = \sum_{Z \subseteq A} f(Z)\delta(Z, A) = f(A)$$

Clearly, $f = F * \mu \iff F = f * \xi$. Indeed, if $f = F * \mu$, then $f * \xi = F * \mu * \xi = F * \delta = F$; if $F = f * \xi$, then $F * \mu = f * \xi * \mu = f * \delta = f$.

For

$$F(A) = P(S \subseteq A) = \sum_{B \subseteq A} f(B) = (f * \xi)(A)$$

we have

$$f(A) = (F * \mu)(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

Similarly, for

$$f(A) = (F * \mu)(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

we get

$$F(A) = (f * \xi)(A) = \sum_{B \subseteq A} f(A)\xi(B, A) = \sum_{B \subseteq A} f(B)$$

This is referred to as *Mobius inversion formula*.

Uncertainty measures, such as probability measures, are defined axiomatically. For belief functions, viewed as uncertainty measures, the axioms (i) and (ii) are natural as far as the intuitive concept of belief is concerned. The axiom (iii), the monotonicity of infinite order, is interesting since it corresponds precisely (and uniquely) to the concept of distribution of random sets. Specifically, let $\mu$ be the Mobius function on the (finite) poset $(2^U, \subseteq)$, and $F : 2^U \to [0, 1]$. Then $F$ is a belief function if and only if $F * \mu$ is a probability density $f$ on $2^U$ with $f(\varnothing) = 0$.

As we will see, in multivariate belief functions, it is convenient to consider Mobius function on product of locally finite partial ordered sets (locally finite posets).

Let $U$ and $V$ be two finite sets, then the product $2^U \times 2^V$ is a (finite) poset with partial order defined by $(A, B) \leq (A', B')$ iff $A \subseteq A'$, and $B \subseteq B'$. If $\mu$ denotes the Mobius function (which is integer-valued and unique) of $(2^U, \subseteq)$, then the Mobius function $\nu$ of $(2^U \times 2^V, \leq)$

$$\nu((A, B), (A', B')) = \mu(A, A')\mu(B, B') = (-1)^{|A' \setminus A| + |B' \setminus B|}$$

*Examples of belief functions*

(a) Here is an example where the domain $U$ (at the place of $\Theta$) can be arbitrary (i.e., finite or not). Let $(U, \mathcal{U}, P)$ be a probability space. The (associated) inner probability is $F : 2^U \to [0, 1]$, defined by

$$F(B) = \sup\{P(A) : A \in \mathcal{U}, A \subseteq B\}$$

We are going to show that $F(.)$ is a belief function on $U$, i.e., satisfying the three axioms (i), (ii), and (iii) of the above axiomatic definition of a belief function.

Clearly $F(\varnothing) = 0$ and $F(U) = 1$. It remains to show that the inner probability $F(.)$ is infinitely monotone (axiom (iii)).

This requires several steps.

(1) For each $B \subseteq U$, the supremum is attained. Indeed, for each $n \geq 1$, let $A_n \in \mathcal{U}$ , $A_n \subseteq B$ be such that

$$P(A_n) \geq F(B) - \frac{1}{n}$$

Let $A = \cup_{n \geq 1} A_n$. Then, $A \in \mathcal{U}$ and $A \subseteq B$, and hence

$$F(B) \geq P(A) \geq P(A_n) \geq F(B) - \frac{1}{n}$$

Since $n$ is arbitrary, it follows that $F(B) = P(A)$.

(2) For any $n \geq 1$, if $B_1, B_2, ..., B_n$ are in $2^U$, and $A_1, A_2, ..., A_n$ are in $\mathcal{U}$ with $A_i \subseteq B_i$, and $F(B_i) = P(A_i)$, $i = 1, 2, ..., n$, then

$$F(\cap_{i=1}^n B_i) = P(\cap_{i=1}^n A_i)$$

This can be seen as follows.

By (1), there is $C \in \mathcal{U}$ , $C \subseteq \cap_{i=1}^n B_i$ such that $F(\cap_{i=1}^n B_i) = P(C)$. Thus it suffices to show that $P(\cap_{i=1}^n A_i) = P(C)$. Without loss of generality, we can assume that $\cap_{i=1}^n A_i \subseteq C$ , since otherwise, replace $C$ by $C \cup (\cap_{i=1}^n A_i)$ by noting that $C \cup (\cap_{i=1}^n A_i) \subseteq (\cap_{i=1}^n B_i)$ and

$$F(\cap_{i=1}^n B_i) = P(C) \leq P(C \cup (\cap_{i=1}^n A_i)) \leq F(\cap_{i=1}^n B_i)$$

so that

$$P(C \cup (\cap_{i=1}^n A_i)) = F(\cap_{i=1}^n B_i)$$

Let $D = C \backslash (\cap_{i=1}^n A_i) = \cup_{i=1}^n (C \cap A_i^c)$. For each $i \in \{1, 2, ..., n\}$, $D \subseteq C \subseteq B_i$ with $D \cup A_i \subseteq B_i$. It follows that $P(D \cup A_i) = F(B_i) = P(A_i)$, and hence $P(D \cap A_i^c) = 0$. Next, since $D \subseteq \cup_{i=1}^n A_i^c$,

$$P(D) = P(D \cap \cup_{i=1}^n A_i^c) = P(\cup_{i=1}^n (D \cap A_i^c)) \leq \sum_{i=1}^n P(D \cap A_i^c) = 0$$

and therefore, $P(\cap_{i=1}^n A_i) = P(C)$.

(3) The inner probability is infinitely monotone. Indeed, let $n \geq 2$, and $B_1, B_2, ..., B_n$ in $2^U$. By (1), there are $A_1, A_2, ..., A_n$ in $\mathcal{U}$ such that $A_i \subseteq B_i$ and $P(A_i) = F(B_i)$, $i = 1, 2, ..., n$.

Since $\cup_{i=1}^n A_i \in \mathcal{U}$ and $\cup_{i=1}^n A_i \subseteq \cup_{i=1}^n B_i$, we have, by (2),

$$F(\cup_{i=1}^n B_i) \geq P(\cup_{i=1}^n A_i) = \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} P(\cap_{i \in I} A_i)$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

(b) Let $P$ be a probability measure on a finite set $U$. Then $P^2(.)$ is a belief function. It suffices to show that $P^2(A) = \sum_{B \subseteq A} f(B)$ where $f(.)$ is a probability density on $2^U$. The following $f(.)$ will do.

$$f(A) = \begin{cases} 0 \quad \text{for...} \ A = \varnothing \quad \text{or} \quad |A| > 2 \\ P^2(x) \quad \text{for} \quad A = \{x\} \\ 2P(x)P(y) \quad \text{for} \quad A = \{x, y\} \end{cases}$$

by noting that $P^2(A) = [\sum_{x \in A} P(x)]^2$.

(c) Let $U = \{1, 5, 10, 20\}$ and $\mathcal{P}$ be the class of probability measures $P$ on $U$ such that

$$P(1) \geq 0.4, \quad P(5) \geq 0.2, \quad P(10) \geq 0.2, \quad P(20) \geq 0.1$$

Then

$$F(A) = \inf\{P(A) : P \in \mathcal{P}\}$$

is a belief function.

Let

$$f(1) = 0.4, \quad f(5) = 0.2, \quad f(10) = 0.2, \quad f(20) = 0.1, \quad f(\{1, 5, 10, 20\}) = 0.1$$

then $f(.)$ is a probability density on $2^U$, and $F(A) = \sum_{B \subseteq A} f(B)$.

## A.3. CAPACITY FUNCTIONALS

To consider **belief functions on infinite domains** in the next part of this lecture**,** it is more convenient to use a dual notion to belief functions, called *capacity functionals*, with is defined as

$$T(A) = 1 - F(A^c)$$

Clearly, $T(\varnothing) = 1 - F(U) = 0$, $T(U) = 1 - F(\varnothing) = 1$, and for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $2^U$,

$$T(\cap_{j=1}^k A_j) = 1 - F(\cup_{j=1}^k A_j^c) \leq 1 - \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i^c)$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

*Axiomatic definition of a capacity functional*

A function $T : 2^U \to [0, 1]$ is called a capacity functional on the finite set $U$, if it satisfies

(i) $T(\varnothing) = 0$

(ii) $T(U) = 1$

(ii) $T$ is *alternating of infinite order, i.e.,*

For any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $2^U$,

$$T(\cap_{j=1}^{k} A_j) \leq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

*Remark*

The associated belief function of a capacity functional $T$ is $F(A) = 1 - T(A^c)$. Now, in view of Theorem 2, there exist a probability space $(\Omega, \mathcal{A}, P)$ and a nonempty random set $S : \Omega \to 2^U$ such that $F(A) = P(S \subseteq A)$. Thus,

$$T(A) = 1 - F(A^c) = 1 - P(S \subseteq A^c) = P(S \cap A \neq \varnothing)$$

In the context of coarse data where $S$ is a coarsening scheme for the (latent) selector $X$, we have $\{X \in A\} \subseteq \{S \cap A \neq \varnothing\}$ so that $P(S \cap A \neq \varnothing) \geq P(X \in A)$ as opposed to $P(S \subseteq A) \leq P(X \in A)$.

*Examples*

(a) Let $\pi : U \to [0, 1]$ and $\alpha : \Omega \to [0, 1]$ be a random variable, uniformly distributed. Consider the random set $S : \Omega \to 2^U$,

$$S(\omega) = \{u \in U : \pi(u) \geq \alpha(\omega)\}$$

The measurable condition for $S$ is easily verified. The capacity functional of $S$ is, for $A \neq \varnothing$,

$$T(A) = P(S \cap A \neq \varnothing) = P(\omega : \max_{u \in A} \pi(u) \geq \alpha(\omega)) = \max_{u \in A} \pi(u)$$

The associated belief function is

$$F(A) = 1 - T(A^c) = 1 - \max_{u \in A^c} \pi(u)$$

and the probability density of $S$ is

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} [1 - \max_{u \in A^c} \pi(u)]$$

Note that the function $\pi(.)$ is precisely the *covering function* of the random set $S$, i.e.,

$$\pi(u) = P(S \cap \{u\} \neq \varnothing\} = P(u \in S)$$

(b) The capacity functional in the previous example is very special: $T(A) = \max_{u \in A} \pi(u)$. As a capacity functional, it is alternating of infinite order. It is *maxitive*, i.e., $T(A \cup B) = \max\{T(A), T(B)\}$ for any $A, B$ in $2^U$. In fact, in this special case, for any index set $I$, $T(\cup_{i \in I} A_i) = \sup_{i \in I} T(A_i)$. In general, a maxitive set function $T$ is monotone increasing and for any $A_1, A_2, ..., A_k$, $T(\cup_{j=1}^k) = \max\{T(A_i) : i = 1, 2, ..., k\}$.

It turns out that maxitive set functions are always alternating of infinite order. Here is the general setting.

**Proposition**

Let $\mathcal{C}$ be a class of subsets of an arbitrary set $\Theta$, containing $\varnothing$ and stable under finite intersections and unions. Let $T : \mathcal{C} \to [0, +\infty)$ be maxitive. Then $T$ is alternating of infinite order.

*Proof*

Note that if $T$ is maxitive then it is monotone increasing since $A \subseteq B$ implies $T(A \cup B) = T(B) = \max\{T(A), T(B)\}$. Clearly, for any $A_1, A_2, ..., A_k$ in $\mathcal{C}$,

$$T(\cup_{j=1}^k A_j) = \max\{T(A_1), T(A_2), ..., T(A_k)\}$$

Without loss of generality, assume $T(A_1) \leq T(A_2) \leq ... \leq T(A_k)$. Let $J = \{1, 2, ..., k\}$.

For $j \in J$, let $J(j)$ be the set of subsets of $\{1, 2, ..., j\}$ that contain $j$. Then $2^J$ is the disjoint union $\cup_{j \in J} J(j)$. Note that

$$\sum_{I \in J(j)} (-1)^{|I|+1} = \begin{cases} 0 & \text{for} \quad j \geq 2 \\ 1 & \text{for} \quad j = 1 \end{cases}$$

Therefore,

$$\sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i) = \sum_{j \in J} \sum_{I \in J(j)} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

$$= \sum_{j \in J} T(A_j) \sum_{I \in J(j)} (-1)^{|I|+1} = T(A_1) \geq T(\cap_{i=1}^k A_i)$$

since any $I \in J(j)$, $j = 1, 2, ..., k$, contains $j$ so that $T(\cup_{i \in I} A_i) = T(A_j)$.

## A.4. COMBINATION OF EVIDENCE

With the viewpoint of random sets, the problem of combination of evidence becomes clear.

Let $X$ and $Y$ be two random sets on $U$ (finite) with (marginal) distributions $F(A) = P(X \subseteq A)$, $G(B) = P(Y \subseteq B)$, respectively. By combination of $X$ and $Y$, it should mean taking their intersection since they are set-valued random elements. Thus, we are interested in the distribution of $X \cap Y$.

If $X$ and $Y$ are independent (as random elements), then their joint density factors:

$$h(A, B) = P(X = A, Y = B) = P(X = A)P(Y = B) = f(A)g(B)$$

and hence the distribution of $X \cap Y$ is determined as follows

$$P(X \cap Y \subseteq A) = \sum_{B \subseteq A} P(X \cap Y = B) =$$

$$\sum_{B \subseteq A} \sum_{C \cap D = B} P(X = C, Y = D) = \sum_{B \subseteq A} \sum_{C \cap D = B} P(X = C)P(Y = D)$$

Now, any belief function on the finite set $U$ (a distribution function of a random set) corresponds to a density on $2^U$ which is zero at $\varnothing$, i.e., the corresponding random set is a **non-empty** random set $X$ ($P(X = \varnothing) = 0$). But while both $X$ and $Y$ are non empty random sets (on the same $U$), the random set $X \cap Y$ might be not, i.e. $P(X \cap Y = \varnothing) > 0$. To remove this situation, we proceed as follows. Let $V = X \cap Y$ has a distribution $S(.)$, and density $s(.)$. Let the range of $V$ be $\{A_1, A_2, ..., A_k, \varnothing\}$. Then $\sum_{j=1}^{k} s(A_j) = 1 - s(\varnothing)$. Let $W$ have the range $\{A_1, A_2, ..., A_k\}$ with density $t(A_j) = \alpha_j s(A_j)$, where the $\alpha_j$ are chosen so that $t(A_j) > 0$ and $\sum_{j=1}^{k} \alpha_j s(A_j) = 1$, resulting in $t(\varnothing) = 0$. A canonical choice is $\alpha_j = \frac{1}{1-s(\varnothing)}$, in which case $t(A_j) = \frac{s(A_j)}{1-s(\varnothing)}$. This leads to consider the distribution function $T(.)$ of a non-empty $X \cap Y = W$ given by

$$T(A) = P(V \subseteq A | V \neq \varnothing) = \begin{cases} \frac{P(\varnothing \neq V \subseteq A)}{P(V \neq \varnothing)} & \text{for} \quad A \neq \varnothing \\ 0 \quad \text{for} \quad A = \varnothing \end{cases}$$

i.e., the "combination of evidence" from two pieces of evidence $X$ and $Y$ (with corresponding belief functions $F$ and $G$) is taken to be the combined evidence (intersection of two random sets) with distribution given by

$$T(A) = (F * G)(A) = P(X \cap Y \subseteq A | X \cap Y \neq \varnothing)$$

When $X$ and $Y$ are taken as independent, this combined distribution (from $F$ and $G$) is referered to as the Dempster's rule combination of evidence.

Clearly, we also need to consider the problem of combining **dependent** pieces of evidence! The following material, mainly drawn from the recent work of Schmelzer (2012), see also Alvarez (2009), provides the preliminary setting for addressing this problem.

**The notion of joint belief functions** *(as multivariate distributions of vectors of random sets)*

Again, for simplicity of notation, we consider the case of two variables. The extension to several variables is straightforward.

For real-valued (point) random variables, there is no difference between a random element with values in $\mathbb{R}^2$ and a vector of two random variables, each with values in $\mathbb{R}$, and hence their "multivariate" and "joint" distributions are the same. The situation is different for random sets.

A *bivariate random set* $S$ is a random element taking values in the power set $2^{U \times V}$ of the product domain $U \times V$, and its probability law is determined by its density function $m : 2^{U \times V} \to [0, 1]$, $m(A) = P(S = A)$, for $A \subseteq U \times V$. The distribution of such a random element is called a *multivariate belief function* in the literature. It is defined as the function, defined on $2^{U \times V}$, $A \to P(S \subseteq A) = \sum_{B \subseteq A} m(B)$, where their "mass" function is $m : 2^{U \times V} \to [0, 1]$, $m(A) = P(S = A)$.

If we have two random sets, $X$ and $Y$, taking values in $2^U$ and $2^V$, respectively, then the vector $(X, Y)$ will take values in $2^U \times 2^V$ and not $2^{U \times V}$. If we identify $(A, B) \in 2^U \times 2^V$ to $A \times B \in 2^{U \times V}$, then $2^U \times 2^V$ is a proper subset of $2^{U \times V}$, consisting of only "rectangles". The distribution of $(X, Y)$ will be referred to as their *joint distribution.* Thus, vectors of random sets are special cases of multivariate random sets. Their joint distribution, on finite domains, is determined by their joint probability density function on $2^U \times 2^V : P(X = A, Y = B)$.

Joint distributions of random set vectors are appropriate for addressing the problem of combination of evidence. It is so, since, as we have two pieces of evidence, represented by two belief functions $F$ and $G$, distributions of two nonempty random sets $X$ and $Y$, respectively, on some common finite domain $U$, we wish to find some "reasonable" joint belief function (a distribution of the nonempty random vector $(X, Y)$) in order to derive the beleif function of the nonempty random set $X \cap Y$ on $U$.

Specifically, the joint belief function (distribution) associated with $(X, Y)$ is just the two-dimensional version of the univariate one, namely, it is a set function $H : 2^U \times 2^V \to [0, 1]$, given by

$$H(A, B) = P(X \subseteq A, Y \subseteq B)$$

First of all, since $2^U \times 2^V$ is finite, the "joint" probability measure $P_{(X,Y)}$ on the power set of $2^U \times 2^V$ is completely determined by its joint density $h : 2^U \times 2^V \to [0, 1]$, where

$$h(A, B) = P(X = A, Y = B)$$

Thus, $H(., .)$ characterizes the probability law of the vector $(X, Y)$.

Now,

$$H(A, B) = P(X \subseteq A, Y \subseteq B) = \sum_{A' \subseteq A, B' \subseteq B} P(X = A, Y = B) = \sum_{A' \subseteq A, B' \subseteq B} h(A', B')$$

the density $h$ is recovered from its distribution $H$ simply by Mobius inversion on the product poset $(2^U \times 2^V, \leq)$ where the partial order $\leq$ on $2^U \times 2^V$ is defined pointwise: $(A', B') \leq (A, B)$ means $A' \subseteq A$ and $B' \subseteq B$. As such, it is well-known that (see any text on Combinatorics) the Mobius function on $(2^U \times 2^V, \leq)$ is simply the product, namely

$$\mu : (2^U \times 2^V) \times (2^U \times 2^V) \to \mathbb{Z}$$

$$\mu[(A', B'), (A, B)] = (-1)^{|A \setminus A'| + |B \setminus B'|}$$

and the Mobius inverse of $H$ is

$$h(A, B) = (F * \mu)(A, B) = \sum_{(A', B') \leq (A, B)} \mu[(A', B'), (A, B)] F(A', B')$$

$$= \sum_{(A',B') \leq (A,B)} (-1)^{|A \setminus A'| + |B \setminus B'|} F(A', B')$$

What are the characteristic properties of a joint distribution $H(A, B) = P(X \subseteq A, Y \subseteq B)$ of a random set vector, assuming that both $X$ and $Y$ are **nonempty** random sets? Note that, to use distributions of random sets to model belief functions, we need the random sets to be nonempty.

Clearly,

(i) $H(\varnothing, \varnothing) = 0$, $H(U, V) = 1$

What is the counterpart of "monotonicity of infinite order" in higher dimensions?

Let $U$ and $V$ be two (nonempty) finite sets. Suppose we wish to consider beliefs on the domain $U \times V$, i.e., **multivariate belief functions.** Then, extending the univariate case, we would consider a function $H : 2^{U \times V} \to [0, 1]$.

As a special case, consider beliefs only on "rectangles" $A \times B \in 2^{U \times V}$. This is the same as considering a **vector of random sets** $(X, Y)$ where $X$ take values in $2^U$ and $Y$ takes values in $2^V$. In other words, $(X, Y) : \Omega \to 2^U \times 2^V$ (viewed as a subset of $2^{U \times V}$ by embedding $(A, B) \in 2^U \times 2^V$ to $A \times B \in 2^{U \times V}$). In this case, and by analogy with the univariate case, the joint distribution of $(X, Y)$, i.e., the function $H : 2^U \times 2^V \to [0, 1]$,

$$H(A, B) = P(X \subseteq A, Y \subseteq B)$$

could be used as a joint belief function, provided that the vector $(X, Y)$ is nonempty, i.e.,

$$P(X = \varnothing, Y = \varnothing) = 0$$

So, first, consider a nonempty random set **vector** $(X, Y)$ and examine its joint distribution.

If for any $k \geq 2$, and $(A_j, B_j) \in 2^U \times 2^V$, $j = 1, 2, ..., k$, we have

$$H(\cup_{j=1}^k A_j, \cup_{j=1}^k B_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} H(\cap_{i \in I} A_i, \cap_{i \in I} B_i)$$

then we say that $H(.,.)$ is **jointly** monotone of infinite order.

Consider the co-product $U \amalg V = (U \times \{1\}) \cup (V \times \{2\})$. Its power set is $\{A \amalg B = (A \times \{1\}) \cup (B \times \{2\} : (A, B) \in 2^U \times 2^V\}$. Identify $(A, B) \in 2^U \times 2^V$ with $A \amalg B$.

Note that the operation $\amalg$ preserves unions and intersections: $\cup_j (A_j \amalg B_j) = (\cup_j A_j) \amalg (\cup_j B_j)$, $\cap_j (A_j \amalg B_j) = (\cap_j A_j) \amalg (\cap_j B_j)$.

**Theorem 3**

Let $(X, Y)$ be a nonempty random set vector. Then its joint distribution $H(A, B) = P(X \subseteq A, Y \subseteq B)$ satisfies the following properties:

(i) $H(\varnothing, \varnothing) = 0$

(ii) $H(U, V) = 1$

(iii) $H(.,.)$ is *jointly* monotone of infinite order.

**Proof**

(i) $H(\varnothing, \varnothing) = P(X = \varnothing, Y = \varnothing) = 0$, by hypothesis.

(ii) Clearly, $H(U, V) = P(X \subseteq U, Y \subseteq V) = 1$

(iii) Define $\Psi : 2^{U \amalg V} \to [0, 1]$ by $\Psi(A \amalg B) = H(A, B)$. Since $X \subseteq A, Y \subseteq B$ is equivalent to $X \amalg Y \subseteq A \amalg B$, we have, by letting $Z = X \amalg Y : \Omega \to 2^{U \amalg V}$

$$\Psi(A \amalg B) = P(Z \subseteq A \amalg B)$$

From the standard result in the univariate case, $\Psi(.)$ is a belief function, noting that $A \amalg B = \varnothing$ if and only if $A = \varnothing$ and $B = \varnothing$ and hence $\Psi(\varnothing) = H(\varnothing, \varnothing) = 0$ by hypothesis.

As such, $\Psi(.)$ is monotone of infinite order, i.e.,

$$\Psi(\cup_{j=1}^k (A_j \amalg B_j)) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} \Psi(\cap_{i \in I} (A_i \amalg B_i))$$

But,

$$\Psi(\cup_{j=1}^k (A_j \amalg B_j)) = \Psi[(\cup_{j=1}^k A_j) \amalg (\cup_{j=1}^k B_j)] = H(\cup_{j=1}^k A_j, \cup_{j=1}^k B_j)$$

and

$$\Psi(\cap_{i \in I} (A_i \amalg B_i) = \Psi[(\cap_{i \in I} A_i) \amalg (\cap_{i \in I} B_i)] = H(\cap_{i \in I} A_i, \cap_{i \in I} B_i)$$

so that

$$H(\cup_{j=1}^{k} A_j, \cup_{j=1}^{k} B_j)) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} H(\cap_{i \in I} A_i, \cap_{i \in I} B_i)$$

*Remarks.*

(1) Note that $H(.,.)$ is monotone of infinite order "in each component" in the sense that, for any fixed $B \subseteq V$, the set function $H(., B) : 2^U \to [0, 1]$ is monotone of infinite order. Indeed,

$$H(A, B) = P(X \subseteq A, Y \subseteq B) = P(X \subseteq A | Y \subseteq B) P(Y \subseteq B)$$

$P(X \subseteq A | Y \subseteq B)$ is the conditional distribution of $X$ given $(Y \subseteq B)$. As such, just like the proof in the univariate case,

$$P(X \subseteq \cup_{j=1}^{k} A_j | Y \subseteq B) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} P(X \subseteq \cap_{i \in I} A_i | Y \subseteq B)$$

Multiplying both sides by $P(Y \subseteq B)$ we get

$$P(X \subseteq \cup_{j=1}^{k} A_j, Y \subseteq B) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} P(X \subseteq \cap_{i \in I} A_i, Y \subseteq B)$$

meaning that $H(.,.)$ is monotone of infinite order in the first component. Similarly, it is also monotone of infinite order in its second component.

The property (iii), namely "jointly monotone of infinite order" is a stronger condition than "monotone of infinite order in each component". The Example 1 below shows that the weaker condition of "monotone of infinite order in each component" is not sufficient to have a random set representation for a joint belief function.

(2) If we insist that the marginal distributions are univariate belief functions, then we have to consider a vector $(X, Y)$ whose components are non-empty random sets. In this case, the condition (i) is obviously satisfied since $P(X = \varnothing, Y = \varnothing) \leq P(X = \varnothing)$. While the Theorem 1 implies that marginal functions $H(., V)$ and $H(U, .)$ are monotone of infinite order, they need not be belief functions unless we assume in addition that $H(\varnothing, V) = H(U, \varnothing) = 0$.

More specifically, if $X$ and $Y$ are nonempty random sets, then the vector $(X, Y)$ is nonempty, and its joint distribution $H(A, B) = P(X \subseteq A, Y \subseteq B)$ is a **joint belief function with belief marginals.**

The upshot is that the above property (i), (ii) and (iii) provide an axiomatic definition of joint belief functions.

As $H$ is a real-valued function , defined on $(2^U \times 2^V, \leq)$, its Mobius inverse is $h = H * \mu$, resulting in $h : 2^U \times 2^V \to [0, 1]$

$$h(A, B) = \sum_{(A',B') \leq (A,B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H(A', B')$$

with

$$H(A, B) = \sum_{(A',B') \leq (A,B)} h(A', B')$$

Thus, to see whether or not a function $H$, satisfying (i), (ii) and (iii) above admits a random set representation, it suffices to find out whether or not $h$ is a bona fide joint probability density on $2^U \times 2^V$ with $h(\varnothing, \varnothing) = 0$.

**Theorem 4**

If $H : 2^U \times 2^V \to [0, 1]$ satisfies

(i) $H(\varnothing, \varnothing) = 0$

(ii) $H(U, V) = 1$

(iii) $H(., .)$ is jointly monotone of infinite order,

then there exist $(\Omega, \mathcal{A}, P)$ and $(X, Y) : \Omega \to 2^U \times 2^V$ , a nonempty random set vector, such that $H(A, B) = P(X \subseteq A, Y \subseteq B)$.

**Proof**

We provide a simple proof using incidence algebras.

The Mobius inverse of $H : 2^U \times 2^V \to [0, 1]$ is $h : 2^U \times 2^V \to [0, 1]$

$$h(A, B) = \sum_{(A',B') \leq (A,B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H(A', B')$$

and hence,

$$H(A, B) = \sum_{(A',B') \leq (A,B)} h(A', B')$$

Thus, it suffices to verify that $h$ is a probability density with $h(\varnothing, \varnothing) = 0$.

In view of the one-to-one correspondence $(A, B) \to A \amalg B$, we can consider the well-defined function $\Psi : 2^{U \amalg V} \to [0, 1]$,

$$\Psi(A \amalg B) = H(A, B)$$

a) $\Psi$ is a belief function on $U \amalg V$.

By (i) and (ii), $\Psi(\varnothing) = H(\varnothing, \varnothing) = 0$, $\Psi(U \amalg V) = H(U, V) = 1$.

For monotonicity of infinite order of $\Psi$, we use the joint monotonicity of infinite order of $H$:

$$\Psi(\cup_{j=1}^{k}(A_j \amalg B_j)) = \Psi((\cup_{j=1}^{k} A_j) \amalg (\cup_{j=1}^{k} B_j)) = H((\cup_{j=1}^{k} A_j), (\cup_{j=1}^{k} B_j))$$

$$\geq \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} H((\cap_{i \in I} A_i), (\cap_{i \in I} B_j)) = \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} \Psi(\cap_{i \in I} A_i) \amalg (\cap_{i \in I} B_j)$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} \Psi(\cap_{i \in I}(A_i \amalg B_j))$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} \Psi((\cap_{i \in I} A_i) \amalg (\cap_{i \in I} B_i))$$

$$= \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} \Psi(\cap_{i \in I}(A_i \amalg B_i))$$

b) From a) and the standard result in the univariate case, it follows that the Mobius inverse of $\Psi$

$$\psi(A \amalg B) = \sum_{A' \amalg B' \subseteq A \amalg B} (-1)^{|(A \amalg B) \setminus (A' \amalg B')|} \Psi(A' \amalg B')$$

is a probability density on $2^{U \amalg V}$.

Now, $A' \amalg B' \subseteq A \amalg B \iff (A', B') \leq (A, B)$, and $|A \amalg B| = |A| + |B|$ which implies that $|(A \amalg B) \setminus (A' \amalg B')| = |A \setminus A'| + |B \setminus B'|$. Thus,

$$\begin{aligned} \psi(A \amalg B) &= \sum_{A' \amalg B' \subseteq A \amalg B} (-1)^{|(A \amalg B) \setminus (A' \amalg B')|} \Psi(A' \amalg B') \\ &= \sum_{(A', B') \leq (A, B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H(A', B') = h(A, B) \end{aligned}$$

Q.E.D.

*Remarks*

(1) In view of Theorem 2, **axiomatically, a joint belief functions** is a set function $H : 2^U \times 2^V \to [0,1]$, satisfying the conditions (i), (ii) and (iii) above.

(2) A two-dimensional **canonical random set representation** $(X, Y)$ for $H(.,.)$ is constructed as follows. First, a "focal elements" of $H$ is a rectangle $A \times B$ such that $h(A, B) \neq 0$, where $h$ is the probability density (by Theorem 2)

$$h(A,B) = \sum_{(A',B') \leq (A,B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H(A', B')$$

Let the focal elements of $H$ be $\{A_i \times B_j : i = 1, 2, ..., n; j = 1, 2, ..., m\}$. Partition the unit square $[0,1]^2$ into $nm$ rectangles $I_i \times J_j$ with $\lambda(I_i \times J_j) = h(A_i, B_j)$, where $\lambda(.)$ denotes the Lebesgue measure on $[0,1]^2$. Let $(\Omega, \mathcal{A}, P) = ([0,1]^2, \mathcal{B}_1 \otimes \mathcal{B}_1, \lambda(.))$. Define $X : \Omega \to 2^U$ by $X(x,y) = A_i$ for $(x,y) \in I_i \times [0,1]$, and $Y : \Omega \to 2^V$ by $Y(x,y) = B_j$ for $(x,y) \in [0,1] \times J_j$. Then $(X,Y) : \Omega \to 2^U \times 2^V$ (identified as $\{A \times B : A \in 2^U, B \in 2^V\}$)

$$(X,Y)(x,y) = (X(x,y), Y(x,y)) = A_i \times B_j \quad \text{for} \quad (x,y) \in I_i \times J_j$$

with

$$P(X = A_i, Y = B_j) = P((X,Y) = A_i \times B_j) = \lambda(I_i \times J_j) = h(A_i, B_j)$$

and

$$H(A,B) = P(X \subseteq A, Y \subseteq B)$$

*Example 1.* Let $U = V = \{0,1\}$ and consider $H : 2^U \times 2^V \to [0,1]$ given by

$$H(A,B) = \begin{cases} 0 & \text{for} \quad |A \times B| = 0 \\ 0.1 & \text{for} \quad |A \times B| = 1 \\ 0.4 & \text{for} \quad |A \times B| = 2 \\ 1 & \text{for} \quad |A \times B| = 4 \end{cases}$$

where $|A|$ denotes the cardinality of the set $A$.

This function $H$ satisfies (i), (ii) and "monotone of infinite order in each component", but **not** "jointly monotone of infinite order".

Indeed, since $|A \times B| = 0$ if and only if either $A$ or $B$ (or both) is empty, we have $H(\varnothing, .) = H(., \varnothing) = 0$. In particular, $H(\varnothing, \varnothing) = 0$. Since $|U \times V| = 4$, $H(U, V) = 1$.

For monotonicity of infinite order in each component, observe that since $H(A, B)$ depends only on the cardinality of $A \times B$, it suffices to check

$$H(U, V) \geq H(U, \{0\}) + H(U, \{1\})$$

and

$$H(U, \{0\}) \geq H(\{0\}, \{0\}) + H(\{1\}, \{0\})$$

which are true.

By Mobius inversion

$$h(A,B) = \sum_{(A',B') \leq (A,B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H(A', B')$$

we get $h(U, V) = -0.2$ so that $h$ is *not* a probability density.

On the other hand, it seems that the investigation of joint belief functions is closely related to the problem of **combination of evidence.**

Recall that random set representations of belief functions in the univariate case provide the firm analysis of rules of combining various pieces of evidence, represented by univariate belief functions. Specifically, let $F$ and $G$ be two belief functions on a common (finite) domain $U$ (with corresponding probability densities $f, g$, respectively). Let $X$ and $Y$ be some random set representations for $F$ and $G$, respectively. By combining the evidence we mean finding the distribution of the random set $X \cap Y$. Now,

$$P(X \cap Y \subseteq A) = \sum_{B \cap C \subseteq A} P(X = B, Y = C)$$

Since the "combined" random set should be nonempty, we consider in fact the conditional distribution $P(X \cap Y \subseteq A | X \cap Y \neq \varnothing)$. Thus, we need a joint probability density

$$h : 2^U \times 2^U \to [0, 1]$$

where $h(A, B) = P(X = A, Y = B)$ from which, the combined density function is

$$(f \otimes g)(A) = \sum_{B \cap C = A} h(B, C)$$

and combined belief function

$$(F \otimes G)(A) = \sum_{B \subseteq A} (f \otimes g)(B) / \sum_{B \cap C \neq \varnothing} h(B, C)$$

Thus, we start out with, say, two belief functions (whose random set representations $X, Y$ are nonempty random sets) and wish to come up with a "joint belief function" for the vector of random sets $(X, Y)$ **admitting the belief functions $F, G$ as marginal belief functions.** In other words, we seek a concept of "joint belief function" whose marginals are belief functions.

As mentioned in the remark following Theorem 3, if $X$ and $Y$ are both nonempty random sets, then necessarily the vector $(X, Y)$ is non empty, but the converse is not true.

*Example 2*
Let $U = V = \{0\}$ and $H : 2^U \times 2^V \to [0, 1]$, defined by
$H(\varnothing, \varnothing) = 0$, $H(U, V) = 1$, and $H(\varnothing, V) = H(U, \varnothing) = 0.5$
Then $H$ satisfies (i), (ii) and (iii) of Theorem 3.
Its Mobius inverse is $h(\varnothing, \varnothing) = h(U, V) = 0$, and $h(\varnothing, V) = h(U, \varnothing) = 0.5$. Thus, the marginal distribution of the random set $X$ is such that $P(X = \varnothing) = P(X = \varnothing, Y \subseteq V) = P(X = \varnothing, Y = V) = h(\varnothing, V) \neq 0$, i.e., $X$ is not an empty random set, and hence it is not a random set representation of a belief function. In other words, marginals of a joint belief function (in the sense of having (i), (ii) and (iii) above) need not be belief functions.

Let $X$ and $Y$ be two canonical random set reprentations of two belief functions $F$ and $G$, on the same finite domain $U$, respectively. Suppose $|\mathcal{F}(F)| = n$ and $|\mathcal{F}(G)| = m$.

Now, the Lebesgue measure $\lambda(dx)$ on $[0,1]$ correponds, by Lebesgue-Stieltjes theorem, to the distribution function $x \rightarrow x$, on $[0,1]$, of the uniform random variable on it. As such, a joint distribution on $[0,1]^2$ with uniform marginals is a **copula**, in the sense of Sklar, i.e., a function $C : [0,1]^2 \rightarrow [0,1]$ such that

a) $C(u,0) = C(0,v) = 0$ for any $u, v \in [0,1]$

b) $C(u,1) = 1$, $C(1,v) = v$ for any $u, v \in [0,1]$

c) For any $u \leq u', v \leq v'$

$$C(u'v') - C(u,v') - C(u',v) + C(u,v) \geq 0$$

meaning simply that $C(.,.)$ is a joint distribution function with support $[0,1]^2$ and having uniform distributions on $[0,1]$ as marginal distribution functions.

As a joint distribution function on $[0,1]^2$, $C$ generates a probability measure on $\mathcal{B}_1 \otimes \mathcal{B}_1$, a $\sigma-$field of $[0,1]^2$, denoted as $dC$, via

$$dC\{[u,u'] \times [v,v']\} = C(u'v') - C(u,v') - C(u',v) + C(u,v)$$

Thus, if we equip the measurable space $([0,1]^2, \mathcal{B}_1 \otimes \mathcal{B}_1)$ with the probability measure $dC$ generated by any (bivariate) copula $C$, for the vector of random sets $(X,Y)$ with values in $2^U \times 2^U$, then the function $H_C : 2^U \times 2^U \rightarrow [0,1]$ , defined by

$$H_C(A,B) = dC\{(x,y) \in [0,1]^2 : X(x) \subseteq A, Y(y) \subseteq B\}$$

$$= dC\{(x,y) \in [0,1]^2 : X(x) \times Y(y) \subseteq A \times B\}$$

is a **joint belief function with marginals** $F$ and $G$.

From

$$H_C(A,B) = dC\{(x,y) \in [0,1]^2 : X(x) \times Y(y) \subseteq A \times B\}$$

we get

$$h_C(A,B) = \sum_{(A',B') \leq (A,B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H_C(A',B')$$

As for the random set of interest in the problem of combination of evidence $X \cap Y$, we take $V = U$. First, we have

$$dC\{X \cap Y = A\} = \sum_{B \cap D = A} dC\{X = B, Y = D\} = \sum_{B \cap D = A} h_C(B, D)$$

which is the density of $X \cap Y$, denoted as $(f \oplus_C g)(A)$. Therefore, the distribution of $X \cap Y$ is

$$(F \oplus G)(A) = \sum_{B \subseteq A} (f \oplus_C g)(B)$$

Finally, to get a nonempty $X \cap Y$ on $2^U$, we normalize the above, just like in the univariate case, leading to the (conditional) probability density on $2^U$ of a "combined" belief for $X \cap Y$ is obtained as

$$(f \oplus_C^* g)(A) = [\sum_{A' \cap B' = A} h_C(A', B')] / [\sum_{A' \cap B' \neq \varnothing} h_C(A', B')]$$

as well as its conditional combined belief function

$$(F \oplus_C^* G)(A) = \sum_{B \subseteq A} (f \oplus_C^* g)(B)$$

which you could call **a rule of combination of evidence when the dependence structure between the evidence is modeled by a copula** $C$.

When the copula $C$ is the product copula $C(x, y) = xy$, it reduces to the well-known "Dempster rule of combination".

Of course, as in any problems involving the choice of a copula, the subtle thing is "how" to pick a copula for a problem at hand?!

**Canonical joint belief functions**

Each canonical random set $X$ of a belief function $F$ is constructed from a given ordering of the focal elements of $F$. If the set of focal elements of $F$, denoted as $\mathcal{F}(F)$, has cardinality $n$, then there are $n!$ canonical random set representations of $X$ which are all equivalent, i.e., having the same distribution.

If $Y$ is a canonical random set representation of another belief function $G$ on $V$, then while for any two canonical random sets $X, X'$ of $F$, and $Y, Y'$ of $G$, we have $X \stackrel{d}{=} X'$, $Y \stackrel{d}{=} Y'$, the pairs $(X, Y)$ and $(X', Y')$ do not have necessarily the same joint distribution. As a result, for each copula $C$ fixed, the joint density of $(X, Y)$, denoted as $h^C_{(X,Y)} : 2^U \times 2^V \to [0, 1]$, differs from

$h^C_{(X',Y')}$, and hence the joint density used in deriving a rule of combination of evidence depends on which pair of canonical random sets we pick (in addition to the copula already picked).

If $\mathcal{F}(G)$ has cardinality $m$, then we have $n!m!$ pairs of canonical random sets!

It turns out that for any given pair of canonical random sets $(X, Y)$, corresponding to a pair of orderings of elements of the focal elements of $f$ and $g$, the class $\{h^C_{(X,Y)}; C \in \mathcal{C}\}$, where $\mathcal{C}$ denotes the set of all copulas, is that of all joint belief densities with given marginals $f, g$, i.e., if $h : 2^U \times 2^V \rightarrow [0,1]$ such that $\sum_{A,B \subseteq U} h(A,B) = 1$ and $f(A) = \sum_{B \subseteq U} h(A,B)$, $g(B) = \sum_{A \subseteq U} h(A,B)$, there, for a given $(X,Y)$, there exists $C \in \mathcal{C}$ such that $h = h^C_{(X,Y)}$. As such, the question is: for a given $(X,Y)$, which copula $C$ to choose to use in combination of evidence?

In the context of evidence, we only have two belief functions $F$ and $G$ (from which we derive their densities $f$ and $g$), and nothing else. Note that the problem of choosing a copula in a practical modeling problem in statistics is in general guided by empirical data, exibiting some dependence, from which various ways to choose an appropriate copula can be concisdered. In the absence of any additional information about the possible dependence between two pieces of evidence, assuming "independence" could be misleading when they are in fact not independent! Thus, one possible way out to calling upon some "acceptable principle", such as the principle of maximum entropy.

Recall that, the entropy of the density $h^C_{(X,Y)}$ is

$$Ent(h^C_{(X,Y)}) = -\sum_{A \subseteq U, B \subseteq V} h^C_{(X,Y)}(A,B) \log h^C_{(X,Y)}(A,B)$$

Thus, to combine $F$ and $G$, we could proceed as follows. Construct a pair $(X,Y)$ from $F$ and $G$, then solve

$$\max_{C \in \mathcal{C}} Ent(h^C_{(X,Y)})$$

to obtain $C^*$ and use $h^{C^*}_{(X,Y)}$ for combining evidence.

# B. BELIEF FUNCTIONS ON INFINITE DOMAINS

**PRELIMINARIES**

The building block for modeling uncertainty in probability theory is the concept of a probability space $(\Omega, \mathcal{A}, P)$ where $\Omega$ is an arbitrary set; the domain $\mathcal{A}$ of the probability measure $P$, as a subset of $2^{\Omega}$, is a $\sigma-$ field, and hence containing $\varnothing$ and $\Omega$; and the set function $P : \mathcal{A} \to [0,1]$ satisfying $P(\varnothing) = 0, P(\Omega) = 1$, and $\sigma-$ additive, implying that, for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $\mathcal{A}$,

$$P(\cup_{j=1}^{k}) = \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} P(\cap_{i \in I} A_i)$$

Weakening this equality leads to the concept of belief functions.

Formally, let $U$ be an arbitrary set (finite ot not). Let $\mathcal{U}$ be a collection of subsets of $U$ ($\mathcal{U} \subseteq 2^U$), containing $\varnothing$ and $U$. A belief function on the domain $U$ is a set function $F : \mathcal{U} \to [0,1]$ satisfying

(i) $F(\varnothing) = 0$

(ii) $F(U) = 1$

(iii) Monotonicity of infinite order: for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $\mathcal{U}$,

$$F(\cup_{j=1}^{k}) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

Note that in the original theory of belief functions (Shafer, 1976), the domain $U$ is a finite set, and $\mathcal{U} = 2^U$, but the above axioms in the finite case do not involve the finiteness of the domain $U$, and hence they make sense for arbitrary domains as well. For example, let $\mathcal{A}$ be a $\sigma-$field of subsets of an arbitrary set $U$, and $P : \mathcal{A} \to [0,1]$ be a probability measure. Let $F : 2^U \to [0,1]$ be defined as, for $B \in 2^U$,

$$F(B) = \sup\{P(A) : A \in \mathcal{A}, \quad A \subseteq B\}$$

Then $F(.)$ is a belief function on $U$ (see the proof in Part A).

In the studies of belief functions, we distinguish various situations concerning their domains. Since we are dealing with set functions (in fact with

random sets), we look at the "range" $\mathcal{U}$ of corresponding random sets (which is the "domain" of belief functions, so that by "belief functions on infinite domains", we mean $\mathcal{U}$ being infinite). As a collection of subsets of $U$, the space $\mathcal{U}$ has a natural partial order relation, namely set inclusion $\subseteq$. From a "combinatorial" viewpoint (for using incidence algebras), there are two main cases: $(\mathcal{U}, \subseteq)$ is locally finite, or not.

For $\mathcal{U} = 2^U$ with $U$ finite, the poset $(\mathcal{U}, \subseteq)$ is locally finite, so that, by Mobius inversion, a belief function $F$ on $U$ is completely characterized by a probability density $f$ on $2^U$ via

$$F(A) = \sum_{B \subseteq A} f(B)$$

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

Thus, most of the researchers took the probability density $f$ as the building thing for belief (univariate and multivariate) functions. This reminds us of the well-known situation in probability theory concerning characterizations of probability laws of random elements: Probability laws of discrete random variables can be chacterized by probability density functions, but for general random variables, that is not the case, and we have to characterize probability laws, say, on euclidian spaces $\mathbb{R}^d$, by their distribution functions which, in turn, are defined axiomatically. The situation is similar to belief functions. On infinite domains $\mathcal{U}$, we have to use the axiomatic definition of belief functions, this is so, since $(\mathcal{U}, \subseteq)$ might not be a locally finite poset. For example, $U = \mathbb{R}$, $\mathcal{U} = \mathcal{B}(\mathbb{R})$.

Now, as noted in Part A, we can study belief functions by looking at their duals called capacity functionals: $T(A) = 1 - F(A^c)$, for which a general axiomatic definition is inspired dually.

On the other hand, when rigorously formulated, the viewpoint of coarse data could provide a "reasonable" concept of degrees of belief in any type of domains, finite or not. In fact, the problem of coarse data arises often in the domain of the reals $\mathbb{R}$: a coarsening $S$ of a selector $X$ takes values as subsets of $\mathbb{R}$, an infinite domain. The "containment" set function $A \to P(S \subseteq A)$ provides an intuitive meaning to the degree of belief of the subset $A$ of $\mathbb{R}$. To use the random set framework as in the case of finite domains, it is necessary to formulate the notion of belief functions on the domain $\mathbb{R}$ in such a way that they correpond to probability laws of random sets on $\mathbb{R}$. In view of this,

and as an tutorial introduction to belief functions on infinite domains, we will spell out a special class of random sets on the real line $\mathbb{R}$ (or on $\mathbb{R}^d$) and investigate their probability laws.

## B.1. RANDOM CLOSED SETS

As for finite random variables, our discussions concerning belief functions on a finite domain $U$ in Part A do not involve topological concepts. The domain of a belief function, or its dual capacity functional, is the whole power set $2^U$ which could be viewed as a topology on $U$. With respect to this discrete topology, not only any subset of $U$ is a closed subset, but also, a "Borel" set, i.e., a member of the $\sigma-$ field generated by the topology. Thus, in this topological context, each finite random set $X$ takes values as *closed* subsets of $U$. Thus, if we wish to consider belief functions on arbitrary domains, finite or infinite, we should consider topological spaces $U$ and random sets taking closed subsets of $U$ as values. Such random sets are called *random closed sets.*

Recall that a topology on a set $U$ is a collection $\mathcal{T}$ of subsets of $U$ having the following properties

a) $\varnothing \in \mathcal{T}$ , $U \in \mathcal{T}$

b) The union of the elements of any subcollection of $\mathcal{T}$ is in $\mathcal{T}$

c) The intersection of the elements of any finite subcollection of $\mathcal{T}$ is in $\mathcal{T}$

The couple $(U, \mathcal{T})$ is called a topological space.The elements of $\mathcal{T}$ are called open sets. Complements of open sets are called closed sets. A subset $A \subseteq U$ is a neighborhood of a point $u \in U$ if it contains an open set containing $u$. $A$ is compact if every open covering of $A$ contains a finite subcollection that also covers $A$. A basis for a topology is a collection $\mathcal{C}$ of subsets of $U$ such that : (1) for each $u \in U$, there is at least one element of $\mathcal{C}$ that contains it, (2) if $u$ belongs to the intersection of two elements of $\mathcal{C}$, then there is an element in $\mathcal{C}$ containing $u$ and contained in that intersection. A basis $\mathcal{C}$ generates a topology $\mathcal{T}(\mathcal{C})$ by saying that a subset $A$ of $U$ is open if for each $x \in A$, there is $C \in \mathcal{C}$ such that $x \in C \subseteq A$. More specifically, $\mathcal{T}(\mathcal{C})$ is the collection of all unions of elements of $\mathcal{C}$.

Moreover, in parallel with the setting of belief functions as "distributions" of random sets on finite domains, we wish to represent belief functions as "distributions" of random closed sets in general.

A nice and practical situation is when the domain $U$ is a locally compact, Hausdorff and second countable (LCHSC) space. This means that, as a

topological space, $U$ is Hausdorff (i.e., for each pair of distinct points in $U$, there exist disjoint neighborhoods of them); locally compact (i.e., every $u \in U$ has a compact neighborhood); second countable (i.e., having a countable basis for its topology). The "canonical" example that we will use is $U = \mathbb{R}^d$.

Clearly, spaces such as $\mathbb{R}^d$ are "practical", but why they are "nice"?

Recall that the "nice" situation for random variables or vectors, i.e., random elements taking values in $\mathbb{R}^d$ is that their probability laws (measures), which are defined on the Borel $\sigma-$field $\mathcal{B}(\mathbb{R}^d)$, are charcterized by simpler objects, namely their distribution functions, which are defined on $\mathbb{R}^d$ (by the Lebesgue-Stieltjes Theorem). This nice situation does not exist for general random elements. However, for random closed sets on LCHSC spaces, there is, fortunately, a counterpart of Lebesgue-Stieltjes Theorem, namely, Choquet Theorem, which allows us to focus on "distributions" of random sets, suitable for our analysis of belief functions. Roughly speaking, Choquet Theorem says that the probability *measure (law)* of a random set can be characterized (determined uniquely) by a set function at a lower level. A form of it already appears in the finite case: the probability measure $P_S$ of a random set $S$ taking values in $2^U$, with $U$ finite, which is defined on the $\sigma-$field $2^{2^U}$, is determined by the capacity functional

$$T : 2^U \to [0,1], \quad T(A) = P(S \cap A \neq \varnothing) = P_S(\mathcal{F}_A)$$

where $\mathcal{F}_A = \{B \in 2^U : A \cap B \neq \varnothing\}$. Note that it is so since $T(A) = 1 - F(A^c)$ where $F(A) = P(S \subseteq A)$ is the distribution (belief function) of $S$, and, by Mobius inversion, it is characterized by the probability density function $f$ on $2^U$: $f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$.

We will denote by $\mathcal{F}, \mathcal{G}$ and $\mathcal{K}$ the classes of closed, open and compact subsets of $\mathbb{R}^d$, respectively.

Let $(\Omega, \mathcal{A}, P)$ be a probability space. To define a random element taking values in the class of *closed sets* $\mathcal{F}$, we need to equip $\mathcal{F}$ with some $\sigma-$field $\mathcal{B}(\mathcal{F})$ to have a measurable space $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ from which, a map $S : \Omega \to \mathcal{F}$ is called a **random cloed set** if it is $\mathcal{A} - \mathcal{B}(\mathcal{F}) -$ measurable, i.e., $S^{-1}[\mathcal{B}(\mathcal{F})] \subseteq \mathcal{A}$.

Following Matheron (1975), we will take $\mathcal{B}(\mathcal{F})$ as the Borel $\sigma-$ field generated by the hit-or-miss topology $\mathcal{T}$ on $\mathcal{F}$ which is constructed as follows.

We denote by $\mathcal{F}, \mathcal{G}$ and $\mathcal{K}$ the classes of closed, open and compact subsets of $\mathbb{R}^d$, respectively. Also, for $A \subseteq \mathbb{R}^d$,

$$\mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \varnothing\}$$

$$\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \varnothing\}$$

$$\mathcal{F}^K_{G_1,G_2,...,G_n} = \mathcal{F}^K \cap \mathcal{F}_{G_1} \cap ... \cap \mathcal{F}_{G_n}$$

Let

$$\mathbb{B} = \{\mathcal{F}^K_{G_1,G_2,...,G_n} : K \in \mathcal{K}, G_i \in \mathcal{G}, i = 1, 2, ..., n\}$$

Note that in $\mathbb{B}$, when $n = 0$, $\mathcal{F}^K_{G_1,G_2,...,G_n}$ stands for $\mathcal{F}^K$. Also, $\mathcal{F}_G = \mathcal{F}^{\varnothing}_G$, $\mathcal{F}_{\varnothing} = \varnothing$, $\mathcal{F} = \mathcal{F}^{\varnothing}$.

The collection $\mathbb{B}$ is a basis for a topology $\mathcal{T}$ which is called the hit-or-miss topology on $\mathcal{F}$. Its Borel $\sigma-$ field is denoted as $\mathcal{B}(\mathcal{F})$.

*Examples*

(1) Let $\varphi : \mathbb{R}^d \to [0, 1]$ be upper semi-continuous (u.s.c.), i.e., $\{x \in \mathbb{R}^d : \varphi(x) \geq a\} \in \mathcal{F}$ for any $a \in \mathbb{R}$. Let $\alpha : \Omega \to [0, 1]$ be a random variable, uniformly distributed on $[0, 1]$. Define $S : \Omega \to \mathcal{F}$ by

$$S(\omega) = \{x \in \mathbb{R}^d : \varphi(x) \geq \alpha(\omega)\}$$

(2) Let $S : \Omega \to \mathcal{F}$ be a random closed on $\mathbb{R}^d$, and $\lambda(dx)$ be the Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$. Then $\lambda(S) : \Omega \to \mathbb{R}^+$ is a random variable. Indeed, let $\psi : \mathbb{R}^d \times \mathcal{F} \to \{0, 1\}$ be

$$\psi(x, F) = \begin{cases} 1 & \text{if} \quad x \in F \\ 0 & \text{if} \quad x \notin F \end{cases}$$

Then the restriction of $\lambda(dx)$ to $\mathcal{F}$ is

$$\lambda(F) = \int_{\mathbb{R}^d} \psi(x, F)\lambda(dx)$$

so that, by Fubini Theorem, $\lambda : \mathcal{F} \to \bar{\mathbb{R}}$ is $\mathcal{B}(\mathcal{F})-$ $\mathcal{B}(\bar{\mathbb{R}})-$ measurable, and hence $\lambda(S) = \lambda \circ S$ is $\mathcal{A} - \mathcal{B}(\bar{\mathbb{R}})-$ measurable.

*Remark*

The example (2) justifies the computation of the expected value of the volume of a random closed set in the well-known Robbins' formula (1944).

As $\lambda(S)$ is a random variable, we can proceed to compute its expectation $E[\lambda(S)]$. It is not easy to obtain the distribution of the random variable $\lambda(S)$ (from the probability law of the random closed set $S$) in order to carry out the computation! The following Robbins' formula, for computing $E[\lambda(S)]$, is interesting since it does not require the knowledge of the distribution of $\lambda(S)$ (or as a matter of fact, the probability law of $S$), but only requires the knowledge of the *covering function of* $S$, i.e., $\pi : \mathbb{R}^d \to [0,1]$, $\pi(x) = P(x \in S)$. It goes as follows.

First, observe that the above function $\psi$ is measurable since

$$\psi^{-1}(\{0\}) = \{(x, F) : x \in \mathbb{R}^d, F \in \mathcal{F}, x \notin F\}$$

$$= \cup_{B \in \mathcal{C}}(B \times \mathcal{F}^B) \in \mathcal{F}(\bar{\mathbb{R}}) \otimes \mathcal{B}(\mathcal{F})$$

where $\mathcal{C}$ is a countable basis for the usual topology of $\mathbb{R}^d$.

As such, by Fubini again,

$$\int_{\mathbb{R}^d \times \mathcal{F}} \psi(s,F) d(\lambda \otimes P_S) = \int_{\mathbb{R}^d} \int_{\mathcal{F}} \psi(s,F) dP_S(F) \lambda(dx) = \int_{\mathbb{R}^d} P(x \in S) \lambda(dx)$$

$$= \int_{\mathcal{F}} \int_{\mathbb{R}^d} \psi(s,F) \lambda(dx) dP_S(F) = \int_{\mathcal{F}} \lambda(F) dP_S(F) = \int_{\mathcal{F}} \lambda(S)(\omega) dP(\omega) = E[\lambda(S)]$$

yielding Robbins' formula

$$E[\lambda(S)] = \int_{\mathbb{R}^d} \pi(x) \lambda(dx)$$

with $\pi(x) = P(x \in S)$.

## B.2. CAPACITY FUNCTIONALS

Since we view belief functions in the context of random sets, it is necessary first to look at probability laws of random sets. Note that the studies of random (closed) sets on euclidean spaces were beautifully achieved by Matheron (1975), from which we borrow results for investigating belief functions on $\mathbb{R}^d$. The counterpart of distribution functions of random variables with values in $\mathbb{R}^d$ is capacity functionals which are "duals" of belief functions. Thus, it suffices to focus on capacity functionals.

Let $S$ be a random closed set on $\mathbb{R}^d$. Its probability law $P_S$ is a probability measure on $\mathcal{B}(\mathcal{F})$, defined by $P_S = PS^{-1}$. We seek a set function defined on some collection of subsets of $\mathbb{R}^d$ which could charcterize $P_S$.

Recall that, in the case of a finite set $U$, the domain of a capacity functional $T$ is the whole power set $2^U$: Let $T : 2^U \to [0, 1]$ be such that

(i) $T(\varnothing) = 0$

(ii) $T$ is alternating of infinite order, i.e., for any $k \geq 2$, and $A_1, A_2, ..., A_k$ in $2^U$,

$$T(\cap_{j=1}^k A_j) \leq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} T(\cup_{i \in I} A_i)$$

Then $T$ characterizes a probability measure $Q$ on $2^{2^U}$, i.e., a probability law of some random set $S$ on $U$, via

$$T(A) = Q\{B \subseteq U, B \cap A \neq \varnothing) = Q(\mathcal{F}_A)$$

More specifically, there exist a probability space $(\Omega, \mathcal{A}, P)$, a random element $S : \Omega \to 2^U$ such that, for any $A \in 2^U$,

$$P(S \cap A \neq \varnothing) = P_S(\mathcal{F}_A) = Q(\mathcal{F}_A) = T(A)$$

In the case of $\mathbb{R}^d$, what should be the domain for corresponding capacity functionals? The following motivates the choice of $\mathcal{K}$.

Random vectors, i.e., random elements taking values in $\mathbb{R}^d$, can be viewed as special random closed sets on $\mathbb{R}^d$ which take singleton sets as values. Specifically, if $X : \Omega \to \mathbb{R}^d$ is a random vector, then $S = \{X\}$ is a random closed set on $\mathbb{R}^d$, and we have, for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_X(A) = P(X \in A) = P(S \cap A \neq \varnothing)$$

However, the probability law $P_X$ of $X$, defined on $\mathcal{B}(\mathbb{R}^d)$, is determined from the values of $P_X$ on the class $\mathcal{K}$ of compact subsets of $\mathbb{R}^d$. Specifically, for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_X(A) = \sup\{P_X(K) : K \in \mathcal{K}, \quad K \subseteq A\}$$

For ease of reference, here is the proof of this fact.

First, $\mathbb{R}^d$ is a metric space, and as such, any probability measure $Q$ on $\mathcal{B}(\mathbb{R}^d)$ is determined by its values on closed or open sets of $\mathbb{R}^d$. This property is called *regularity:* for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$Q(A) = \sup\{Q(F) : F \in \mathcal{F},\ F \subseteq A\} = \inf\{Q(G) : G \in \mathcal{G},\ G \supseteq A\}$$

The regularity of probability measures on metric spaces can be proved simply as follows.

Let $\mathcal{D}$ denote the class of sets $A \in \mathcal{B}(\mathbb{R}^d)$ such that

$$Q(A) = \sup\{Q(F) : F \in \mathcal{F},\ F \subseteq A\} = \inf\{Q(G) : G \in \mathcal{G},\ G \supseteq A\}$$

Since $\mathbb{R}^d \in \mathcal{D}$, $\mathcal{D} \neq \varnothing$. Clearly, $\mathcal{D}$ is closed under complementation: $A \in \mathcal{D} \Longrightarrow A^c \in \mathcal{D}$.

Note that, the regularity property of a set $A$ is equivalent to: for any $\varepsilon > 0$, there exist $F_\varepsilon \in \mathcal{F}$, and $G_\varepsilon \in \mathcal{G}$ such that $F_\varepsilon \subseteq A \subseteq G_\varepsilon$ and $Q(G_\varepsilon \backslash F_\varepsilon) < \varepsilon$.

Let $\varepsilon > 0$, $A_n \in \mathcal{D}$, and $F_n \in \mathcal{F}, G_n \in \mathcal{G}$ such that $F_n \subseteq A_n \subseteq G_n$ with $Q(G_n \backslash F_n) < \frac{\varepsilon}{2^{n+1}}$.

Since $\cup_{n \leq N} F_n \nearrow \cup_{n \geq 1} F_n$ as $N \to \infty$, it follows by the continuity of $Q$ that there exists $N$ such that

$$Q[(\cup_{n \geq 1} F_n) \backslash (\cup_{n \leq N} F_n)] < \frac{\varepsilon}{2}$$

For $F = \cup_{n \leq N} F_n \in \mathcal{F}$, we have $F \subseteq \cup_{n \geq 1} A_n \subseteq \cup_{n \geq 1} G_n = G \in \mathcal{G}$, and

$$Q[(\cup_{n \geq 1} G_n) \backslash (\cup_{n \leq N} F_n)] \leq \sum_{n \geq 1} Q(G_n \backslash F_n) + Q[(\cup_{n \geq 1} F_n) \backslash (\cup_{n \leq N} F_n)] < \varepsilon$$

Thus, $\cup_{n \geq 1} A_n \in \mathcal{D}$ and hence $\mathcal{D}$ is a $\sigma-$field.

On the other hand, $\mathcal{F} \subseteq \mathcal{D}$. Indeed, let $F \in \mathcal{F}$. Then $G_n = \{x : \rho(x, F) < \frac{1}{n}\}$ is open and $\searrow F$ since the map $x \to \rho(x, A)$ is continuous (where for a metric $\rho$ on $\mathbb{R}^d, \rho(x, F) = \inf\{\rho(x, y) : y \in F\}$). Therefore, there exists $N$ such that $Q(G_N \backslash A) < \varepsilon$, noting that $F \subseteq G_N$. Thus, $\mathcal{D} = \mathcal{B}(\mathbb{R}^d)$.

Now, since the metric space $\mathbb{R}^d$ is complete and separable, probability measures on $\mathcal{B}(\mathbb{R}^d)$ are determined by their values on compact sets. This is a consequence of the fact that such probability measures are *tight*, i.e., if $Q$ is a probability measure on $\mathcal{B}(\mathbb{R}^d)$, then for each $\varepsilon > 0$, there exists $K_\varepsilon \in \mathcal{K}$ such that $Q(K_\varepsilon) > 1 - \varepsilon$. Indeed, let $\varepsilon > 0>$ For each $n \geq 1$, let $S_n(x) = \{y \in \mathbb{R}^d : \rho(x, y) < \frac{1}{n}\}$. Then the family $\{S_n(x) : x \in \mathbb{R}^d\}$ is a open

covering of $\mathbb{R}^d$, and since $\mathbb{R}^d$ is separable, there is a countable subcovering, say, $\{S_n(x_j) : j \geq 1\}$, such that

$$\mathbb{R}^d = \cup_{j\geq 1} S_n(x_j) = \cup_{j\geq 1} \bar{S}_n(x_j)$$

Now, the sequence $\cup_{j\leq m} \bar{S}_n(x_j) \in \mathcal{F} \nearrow \mathbb{R}^d$ as $m \to \infty$ so that there is $j_o$ such that

$$Q[\cup_{j\leq j_o} \bar{S}_n(x_j)] > 1 - \frac{\varepsilon}{2^n}$$

Let $K_\varepsilon = \cap_{n\geq 1} B_n$ where $B_n = \cup_{j\leq j_o} \bar{S}_n(x_j)$. Then, for each $n$, $K_\varepsilon \subseteq B_n$, and hence $K_\varepsilon$ is compact (a topological fact), and

$$Q(K_\varepsilon^c) \leq \sum_{n\geq 1} Q(B_n^c) < \varepsilon$$

Now, for any $A \in \mathcal{B}(\mathbb{R}^d)$ and $\varepsilon > 0$, since $Q$ is regular, there is $F_\varepsilon \in \mathcal{F}$ such that $F_\varepsilon \subseteq A$, such that

$$Q(A) - Q(F_\varepsilon) = Q(A \backslash F_\varepsilon) < \frac{\varepsilon}{2}$$

Since $Q$ is tight, for each $n$, let $K_n \in \mathcal{K}$ such that $Q(K_n^c) \leq \frac{1}{n}$ so that $Q(K_m^c) < \frac{\varepsilon}{2}$ for $m > [\frac{2}{\varepsilon}]$. Take $K_\varepsilon = F_\varepsilon \cap K_m \in \mathcal{K}$. Then, $K_\varepsilon \subseteq F_\varepsilon \subseteq A$ and $Q(A \backslash K_\varepsilon) < \varepsilon$, i.e.,

$$Q(A) = \sup\{Q(K) : K \subseteq A, K \in \mathcal{K}\}$$

From the above fact, we see that it suffices to consider $P_X$ on $\mathcal{K} \subseteq \mathcal{B}(\mathbb{R}^d)$.

If we let $T : \mathcal{K} \to [0, 1]$ where

$$T(K) = P(X \in K) = P_X(K)$$

then the set function $T(.)$ determines completely the probability law $P_X$ of the random vector $X$. It can be verified that, $T(.)$ has the following properties

(i) $T(\varnothing) = 0$

(ii) If $K_n \in \mathcal{K}$ such that $K_n \searrow K$ (i.e., $K_n \subseteq K_{n+1}$, $K = \cap_{n\geq 1} K_n$), then $T(K_n) \searrow T(K)$

(iii) $T$ is alternating of infinite order, i.e., for $n \geq 2$, and $K_1, K_2, ..., K_n$ in $\mathcal{K}$,

$$T(\cap_{j=1}^{n} K_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} T(\cup_{i \in I} K_i)$$

Now let $S = \{X\}$ be a singleton-valued random closed set on $\mathbb{R}^d$, we have

$$P_X(K) = P(X \in K) = P(\{X\} \cap K \neq \varnothing) = P(S \cap K \neq \varnothing)$$

Thus, for a general random closed set $S$ on $\mathbb{R}^d$, we consider the set function $T_S : \mathcal{K} \to [0,1]$ defined by

$$T_S(K) = P(S \cap K \neq \varnothing) = P_S(\mathcal{F}_K)$$

*Definitions*

(1) The set function $T_S : \mathcal{K} \to [0,1]$ defined by

$$T_S(K) = P(\omega : S(\omega) \cap K \neq \varnothing) = P_S(\mathcal{F}_K)$$

is called the **capacity functional** of $S$.

(2) A set function $T : \mathcal{K} \to [0,1]$ such that

(i) $T(\varnothing) = 0$

(ii) If $K_n \in \mathcal{K}$ such that $K_n \searrow K$ (i.e., $K_n \subseteq K_{n+1}$, $K = \cap_{n \geq 1} K_n$), then $T(K_n) \searrow T(K)$

(iii) $T$ is alternating of infinite order, i.e., for $n \geq 2$, and $K_1, K_2, ..., K_n$ in $\mathcal{K}$,

$$T(\cap_{j=1}^{n} K_j) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} T(\cup_{i \in I} K_i)$$

is called a *capacity functional.*

Just like the special case of $S = \{X\}$, the capacity functional $T_S$ of any random closed set $S$ on $\mathbb{R}^d$ satisfies the above three basic properties (i),(ii) and (iii).

The upshot is that these properties characterize probability laws of randon closed sets. For proofs, see Matheron (1975) or Nguyen (2006).

**Choquet Theorem**

If $T : \mathcal{K} \to [0,1]$ is a capacity functional, then $T$ is the capacity functional of some random closed set $S$ on $\mathbb{R}^d$. Specifically, there exists a unique probability measure $Q$ on $\mathcal{B}(\mathcal{F})$ such that, for all $K \in \mathcal{K}$, $Q(\mathcal{F}_K) = T(K)$.

*Remark*

(1) The capacity functional $T$, defined on $\mathcal{K}$, can be extended to the power set of $\mathbb{R}^d$, preserving the above characteristic properties, as follows.

For $G \in \mathcal{G}$, set

$$T(G) = \sup\{T(K) : K \subseteq G, K \in \mathcal{K}\}$$

then, for $A \subseteq \mathbb{R}^d$,

$$T(A) = \inf\{T(G) : G \supseteq A, G \in \mathcal{G}\}$$

(2) The dual of $T$ is referred to as a belief function on $\mathbb{R}^d$: $A \to 1 - T(A^c)$.

*Example*

Let $\varphi : \mathbb{R}^d \to [0,1]$ be u.s.c. and $T(K) = \sup_{x \in K} \varphi(x)$. Then, $T(.)$ is a capacity functional.

While a probabilistic proof is simple, we verify the basic properties of $T(.)$ in this example directly.

Since the supremum over the empty set is zero, $T(\varnothing) = 0$. $T(.)$ is alternating of infinite order since it is a maxitive set function. Finally, the continuity from above of $T(.)$ is seen as follows.

Let $K_n \searrow K$ in $\mathcal{K}$. We have

$$T(K) \leq \inf_n T(K_n) = s$$

Let $\varepsilon > 0$ and consider

$$A_n = \{x \in \mathbb{R}^d : \varphi(x) \geq s - \varepsilon\} \cap K_n$$

By hypothesis, the set $\{x \in \mathbb{R}^d : \varphi(x) \geq s - \varepsilon\}$ is closed, and so is $A_n$. The $A_n$ are contained in $K_1$, and hence $\cap_{n \geq 1} A_n \neq \varnothing$ (since the $A_n$ have the finite intersection property, and noting that $A_n \neq \varnothing$ for each $n$). Since $A_n \subseteq K_n$ for all $n$, it follows that

$$T(K) \geq \sup_{x \in \cap_{n \geq 1} A_n} \varphi(x) \geq s - \varepsilon$$

implying that $\lim_{n \to \infty} T(K_n) = T(K)$.

*Joint capacity functionals*

Just like the case of finite domains, we consider the setting of joint distributions of random closed set *vectors* in order to investigate the problem of combination of evidence on infinite domains. The recent work of Schmelzer (2012) seems to be the best presentation of the issue.

Again, for simplicity, we address the case of two random closed sets $X$ and $Y$ on the same domain $\mathbb{R}^d$. They are defined on a probability space $(\Omega, \mathcal{A}, P)$ with values in $\mathcal{F}(\mathbb{R}^d)$. We consider the vector $(X, Y) : \Omega \to \mathcal{F} \times \mathcal{F} : (X, Y)(\omega) = X(\omega) \times Y(\omega)$.

The $\sigma-$field on $\mathcal{F} \times \mathcal{F}$ is $\mathcal{B}(\mathcal{F}) \otimes \mathcal{B}(\mathcal{F})$ which is generated by sets $\mathcal{F}_K \times \mathcal{F}_{K'}, K, K' \in \mathcal{K}$.

If we let $P_{(X,Y)}$ be the probability law of $(X, Y)$ on $\mathcal{B}(\mathcal{F}) \otimes \mathcal{B}(\mathcal{F})$, then

$$\begin{aligned} P_{(X,Y)}(\mathcal{F}_K \times \mathcal{F}_{K'}) &= P((X \times Y) \cap (K \times K') \neq \varnothing) \\ &= P(X \cap K \neq \varnothing, Y \cap K' \neq \varnothing) = P(X^-(K) \cap Y^-(K')) \end{aligned}$$

where we set $X^-(K) = \{\omega : X(\omega) \cap K \neq \varnothing\}$.

As in the case of finite domains, we identify each $(K, K')$ with their co-product $K \amalg K' = (K \times \{1\}) \cup (K' \times \{2\})$ which is a subset of $\mathcal{K} \amalg \mathcal{K}$. Similarly, we identify $\mathcal{F} \times \mathcal{F}$ with their co-product $\mathcal{F} \amalg \mathcal{F} = (\mathcal{F} \times \{1\}) \cup (\mathcal{F} \times \{2\})$. In this way, the vector $(X, Y)$ is identified with the random set $S(\omega) = X(\omega) \amalg Y(\omega)$ on $\mathcal{F} \amalg \mathcal{F}$. Note that

$$\{(X \amalg Y) \cap (K \amalg K') \neq \varnothing\} = \{X^-(K) \cup Y^-(K')\}$$

Thus, the (**multivariate**) capacity functional of $(X, Y)$ can be written as

$$T_{(X,Y)} : \mathcal{K} \times \mathcal{K} \to [0, 1] : T_{(X,Y)}(K, K') = P(X^-(K) \cup Y^-(K'))$$

It can be verified that

(i) $T_{(X,Y)}(\varnothing, \varnothing) = 0$

(ii) jointly alternating of infinite order, i.e., for any $k \geq 2$, and $K_j, K'_j, j = 1, 2, ..., k$ compacts,

$$T(\cap_{j=1}^k K_j, \cap_{j=1}^k K'_j) \leq \sum_{\varnothing \neq I \subseteq \{1,2,...,k\}} (-1)^{|I|+1} T(\cup_{i \in I} K_i, \cup_{i \in I} K'_i)$$

(iii) jointly continuous from above, i.e., for any decreasing sequences of compacts $K_n, K'_n$, we have

$$\lim_{n\to\infty} T(K_n, K'_n) = T(\cap_{n\geq 1} K_n, \cap_{n\geq 1} K'_n)$$

As expected, a multivariate version of Choquet Theorem is this.

Given a multivariate capacity functional $T$, defined on $\mathcal{K}\times\mathcal{K}$, we define the associated set function on $\mathcal{K}\amalg\mathcal{K}$ by $T^*(K\amalg K') = T(K, K')$. Then, $T^*$ is a (univariate) capacity functional if and only if $T$ is a multivariate capacity funtional.

*Multivariate Choquet Theorem* (Schmelzer, 2012)

Let $T$ be a multivariate capacity functional on $\mathcal{K}\times\mathcal{K}$. Then there exists a unique probability measure $Q$ on $\mathcal{B}(\mathcal{F}\times\mathcal{F})$ such that $T(K, K') = Q(\mathcal{F}^*_K \cup \mathcal{F}^*_{K'})$, for all $K, K'$ in $\mathcal{K}$ and where $\mathcal{F}^*_K = \{(F, F') \in \mathcal{F}\times\mathcal{F} : F\cap K \neq \varnothing\}$, $\mathcal{F}^*_{K'} = \{(F, F') \in \mathcal{F}\times\mathcal{F} : F'\cap K' \neq \varnothing\}$.

Equiped with joint capacity functionals of random closed sets (or, dually, with joint belief functions) as developped above, research efforts could focus on how to obtain joint belief functions from the knowledge of marginal belief functions, and how to derive belief functions correponding to intersections of nonempty random closed sets. This seems to be a challenging research problem!

# REFERENCES


[1] Alvarez, D.A. (2009), A Momte Carlo-based method for the estimation of lower and upper probabilities of events using infinite random sets of indexable type, *Fussy Sets and Systems* (160), 384-401

[2] Dempster, A. (1967) Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist. (38) 325-339*

[3] Durante, F. and Spizzichino, F. (2010) Simi-copulas, capacities and families of level sets, *Fussy Sets and Systems (*161) 269-276

[4] Matheron, G. (1975) *Random sets and Integral Geometry*, J. Wiley

[5] Nelson, R.B. (1999) *An Introduction to Copulas,* Lecture Notes in Statistics #139, Springer-Verlag

[6] Nguyen, H.T. (1978) On random sets and belief functions, *J. Math. Anal. and Appl. (*65) 531-542

[7] Nguyen, H.T. (2006) *An Introduction to Random Sets,* Chapman and Hall/CRC

[8] Rota, G.C. (1964) On the foundations of combinatorial theory I: Theory of Mobius functions, *A.Wahrsch. (2) 340-368*

[9] Scarsini, M. (1996) Copulae of capacities on product spaces, in *Distributions with Fixed Marginals and Related Topics,* IMS Lecture Notes, Monograph Series, vol. 28, 307-318

[10] Schmelzer, B. (2012) Characterizing joint distributions of random sets by multivariate capacities, *Intern. J. Approximate Reasoning (*53) 1228-1247

[11] Shafer, G. (1976) *A Mathematical Theory of Evidence,* Princeton University Press